BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT/AU98/00817
09/509482

Patent Office
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PO 9500 for a patent by GARVAN INSTITUTE OF MEDICAL RESEARCH filed on 29 September 1997.

I further certify that the annexed specification is not, as yet, open to public inspection.

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

WITNESS my hand this Fifteenth day of October 1998

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

GARVAN INSTITUTE OF MEDICAL RESEARCH

PROVISIONAL SPECIFICATION

Invention Title:

hVDR Isoforms

The invention is described in the following statement:

Field of the Invention:-

The present invention relates to isolated polynucleotide molecules which encode novel isoforms of the human Vitamin D receptor (hVDR). In addition, the present invention relates to the use of these polynucleotide molecules in the production of VDR isoforms using recombinant technology. The polynucleotide molecules and VDR isoforms may be utilised in methods of screening compounds for agonists and/or antagonists.

Background of the Invention:-

The active hormonal form of vitamin D, 1,25-dihydroxyvitamin D3 (1,25(OH)2D3), has a central role in calcium and phosphate homeostasis, and the maintenance of bone. Apart from these calcitropic effects, 1,25-(OH)2D3 has been shown to play a role in controlling cell growth and differentiation in many target tissues. The effects of 1,25-(OH)2D3 are mediated by a specific receptor protein, the vitamin D receptor (VDR), a member of the nuclear receptor superfamily of transcriptional regulators which also includes steroid, thyroid and retinoid receptors as well as a growing number of orphan receptors. Upon binding hormone the VDR regulates gene expression by direct interaction with specific sequence elements in the promotor regions of hormone responsive target genes. This transactivation or repression involves multiple interactions with other protein cofactors, heterodimerisation partners and the transcription machinery.

Although a cDNA encoding the human VDR was cloned in 1988 (1), little has been documented characterising the gene structure and pattern of transcription since that time. The regulation of VDR abundance is one potentially important mechanism for modulating 1,25-(OH)2D3 responsiveness in target cells. It is also possible that VDR has a role in non-transcriptional pathways, perhaps via localization to a non-nuclear compartment and/or interaction with components of other signalling pathways. However, the question of how VDRs are targetted to different cell types and how they are regulated remains unresolved. There have been many reports in the literature describing translational or transcriptional control of VDR levels, both homologously and heterologously, mostly in non-human systems.

A recent study (2) showed that in the kidney, alternative splicing of human VDR transcripts transcribed from a GC rich promotor generates

several transcripts which vary only in their 5' UTRs. The present inventors have now identified further upstream exons of the VDR gene which generate 5' variant transcripts, suggesting that the expression of the VDR gene is regulated by more than one promotor. A subset of these transcripts is expressed in a restricted tissue-specific pattern and further variant transcripts have the potential to encode an N-terminally variant protein. These results may have implications for understanding the actions of 1,25-(OH)2D3 in different tissues and cell types, and the possibility that N-terminally variant VDR proteins may be produced has implications for altered activities such as transactivation function or subcellular localisation of the receptor protein. Furthermore, these variants, by their level, tissue specificity, subcellular localisation and functional activity, may yield targets for pharmaceutical intervention. The variants may also be useful in screening potential analogs and/or antagonists of vitamin D compounds.

Disclosure of the Invention:-

In a first aspect, the invention provides an isolated polynucleotide molecule encoding a human Vitamin D receptor (hVDR) isoform, said polynucleotide molecule comprising a nucleotide sequence which includes sequence that substantially corresponds or is functionally equivalent to that of exon 1b of the human VDR gene.

Exon 1b is a 96 bp exon located 296 bp downstream from exon 1a (5). The sequence of exon 1b is:
5'GTTTCCTTCTTCTGTCGGGGCGCCTTGGCATGGAGTGGAGGAATAAGAA AAGGAGCGATTGGCTGTCGATGGTGCTCAGAACTGCTGGAGTGGAGG3'.

The nucleotide sequence of the polynucleotide molecule of the first aspect of the invention, preferably does not include sequence corresponding to that of exon 1a, exon 1e(5') and/or exon 1e(3'). However, the nucleotide sequence polynucleotide may or may not include sequence corresponding to that of exon 1c and/or exon 1.

Preferably, the polynucleotide molecule of the first aspect comprises a nucleotide sequence which includes;

(i) sequence that substantially corresponds or is functionally equivalent to that of exons 1b and 1-9 and encodes a VDR isoform of approximately 477 amino acids,

(ii) sequence that substantially corresponds or is functionally equivalent to that of exons 1b and 2-9 and encodes a VDR isoform of approximately 450 amino acids, or

(iii) sequence that substantially corresponds or is functionally equivalent to that of exons 1b and 2-9 and further includes a 152 bp intronic sequence substantially corresponding to that shown in Figure 6, and encodes a truncated VDR isoform of approximately 72 amino acids.

Most preferably, the polynucleotide molecule of the first aspect of the invention comprises a nucleotide sequence substantially corresponding or functionally equivalent to that shown in Figure 4, Figure 5, or Figure 6.

The polynucleotide molecule may be incorporated into plasmids or expression vectors (including viral vectors), which may then be introduced into suitable bacterial, yeast, insect and mammalian host cells. Such host cells may be used to express the VDR isoform encoded by the isolated polynucleotide molecule.

Accordingly, in a second aspect, the present invention provides a mammalian, insect, yeast or bacterial host cell transformed with the polynucleotide molecule of the first aspect.

In a third aspect, the present invention provides a method of producing a VDR isoform or a functionally equivalent fragment thereof, comprising culturing the host cell of the second aspect under conditions enabling the expression of the polynucleotide molecule and, optionally, recovering the VDR isoform or fragment thereof.

Preferably, the host cell is of mammalian origin. Preferred examples include Chinese hamster ovary (CHO) cell and the human embryonic kidney cell 293.

In a preferred embodiment, the VDR isoform or fragment thereof is localised to a cell membrane or other subcellular compartment as distinct from a nuclear localisation.

The polynucleotide molecules of the first aspect of the invention encode novel VDR isoforms which may be of interest both clinically and commercially. By using the polynucleotide molecules of the present invention it is possible to obtain VDR isoform proteins or functionally equivalent fragments thereof in a substantially pure form.

Accordingly, in a fourth aspect, the present invention provides a VDR isoform or functionally equivalent fragment thereof encoded by a

polynucleotide molecule of the first aspect, said VDR isoform being in a substantially pure form.

In a fifth aspect, the present invention provides an antibody or antibody fragment capable of specifically binding to the VDR isoform of the fourth aspect.

In a sixth aspect, the present invention provides a non-human animal transformed with a polynucleotide molecule according to the first aspect of the invention.

In a seventh aspect, the invention provides a method for detecting agonist and/or antagonist compounds of a VDR isoform of the fourth aspect, comprising contacting said VDR isoform, functionally equivalent fragment thereof or a cell transformed with and expressing the polynucleotide molecule of the first aspect, with a test compound under conditions enabling the activation of the VDR isoform or functionally equivalent fragment thereof, and detecting an increase or decrease in the activity of the VDR isoform or functionally equivalent fragment thereof.

In an eighth aspect, the present invention provides an oligonucleotide or polynucleotide probe comprising a nucleotide sequence of 10 or more nucleotides capable of specifically hybridising to a unique sequence within the polynucleotide molecule of the first aspect.

In a ninth aspect, the present invention provides an antisense polynucleotide molecule comprising a nucleotide sequence capable of specifically hybridising to an mRNA molecule which encodes a VDR isoform so as to prevent translation of the mRNA molecule.

Such antisense polynucleotide molecules may include a ribozyme region to catalytically inactivate mRNA to which it is hybridised.

The polynucleotide molecule of the first aspect of the invention may be a dominant negative mutant which encodes a gene product causing an altered phenotype by, for example, reducing or eliminating the activity of endogenous VDR isoforms.

In a tenth aspect, the invention provides an isolated polynucleotide molecule comprising a nucleotide sequence substantially corresponding or, at least, showing >75% (preferably >90% or, even more preferably, >95%) identity to:

(i) TGCGACCTTGGCGGTGAGCCTGGGGACAGGGGTGAGGCCAGAGA

CGGACGGACGCAGGGGCCCGGCCCAAGGCGAGGGAGAACAGCGGCACTA
AGGCAGAAAGGAAGAGGGCGGTGTGTTCACCCGCAGCCCAATCCATCAC
TCAGCAACTCCTAGACGCTGGTAGAAAGTTCCTCCGAGGAGCCTGCCATC
CAGTCGTGCGTGCAG (exon 1e (5')),

(ii) AGGCAGCATGAAACAGTGGGATGTGCAGAGAGAAGATCTGGGTC
CAGTAGCTCTGACACTCCTCAGCTGTAGAAACCTTGACAACTCTGCACAT
CAGTTGTACAATGGAACGGTATTTTTTACTCTTCATGTCTGAAAAGGCTA
TGATAAAGATCAA (exon 1e(3')),

(iii) GTTTCCTTCTTCTGTCGGGGCGCCTTGGCATGGAGTGGAGGAATA
AGAAAAGGAGCGATTGGCTGTCGATGGTGCTCAGAACTGCTGGAGTGGA
GG (exon 1b).

The polynucleotide molecules of the tenth aspect may be useful as probes for the detection of VDR variant transcripts and as such may be useful in assessing cell or tissue-specific expression of variant transcripts.

The terms "substantially corresponds" and "substantially corresponding" as used herein in relation to nucleotide sequences is intended to encompass minor variations in the nucleotide sequence which due to degeneracy in the DNA code do not result in a substantial change in the encoded protein. Further, this term is intended to encompass other minor variations in the sequence which may be required to enhance expression in a particular system but in which the variations do not result in a decrease in biological activity of the encoded protein.

The terms "functionally equivalent" as used herein in relation to nucleotide sequences encoding a VDR isoform is intended to encompass nucleotide sequence variants of up to 5% sequence divergence (i.e. retaining 95% or more sequence identity) which encode VDR isoforms of substantially equivalent biological activity(ies) as said VDR isoform.

The term "functionally equivalent fragment" as used herein in respect of a VDR isoform is intended to encompass functional peptide and polypeptide fragments of said VDR isoform which include the domain or domains which bestow the biological activity characteristic of said VDR isoform.

The terms "comprise", "comprises" and "comprising" as used throughout the specification are intended to refer to the inclusion of a stated

component or feature or group of components or features with or without the inclusion of a further component or feature or group of components or features.

The invention will hereinafter be further described by way of the following non-limiting example and accompanying figures.

Brief description of the figures:-

Figure 1(A) provides a map of the human vitamin D receptor (hVDR) gene showing the exon/intron structure, (B) provides the structures of hVDR transcripts: transcripts 1-4 originate from exon 1a [transcript 1 corresponds to the published cDNA sequence (1)], transcripts 5-8 originate from exon 1b and transcripts 9-12 originate from exon 1e(5').

Figure 2 provide the results of RT-PCR analysis of hVDR gene expression: (A) exon 1a transcripts, (B) exon 1b transcripts, and (C) exon 1e(5') transcripts. Lanes 1 and 18 are molecular weight markers (PUC 19 restricted with HpaII). Lane 2-kidney, lane 3-parathyroid adenoma, lane 4-LIM 1863, lane 5-placenta, lane 6-osteoclastoma, lane 7-leukocytes, lane 8-BC1, lane 9-MG-63, lane 10-Saos-2, lane 11-HK-2, lane 12-HEK-293, lane 13-Intestine-407, lane 14-COLO 206F, lane 15-T47D, lane-16WS1, lane 17-no cDNA control.

Figure 3 provides the nucleotide sequence of novel exons detected by 5' RACE: (A) exon 1c, (B) exon 1e(5') [P1e(5') is indicated by an arrow above the sequence], (C) exon 1e(3'), (D) exon 1b [in-frame ATG condons are highlighted and P1b is indicated by an arrow above the sequence]. Intronic sequences are shown in lower case. Canonical splice site consensus sequences are indicated in bold. The transcription start sites for exons 1e(5') and 1b were determined by the 5' termini of RACE clones. No intron sequence is shown 3' to exon 1e(5') as cosmid clone J5 terminated in the intron between exons 1e(5') and 1e(3').

Example:-

EXPERIMENTAL PROCEDURES

Isolation and Characterisation of Genomic Clones

A human lymphocyte cosmic library (Stratagene, La Jolla, Ca) was screened using a 2.1kb fragment of the hVDR cDNA encompassing the entire coding region but lacking the 3'UTR, a 241 bp PCR product spanning exons 1 to 3 of the human VDR cDNA, and a 303 bp PCR product spanning exons 3 and 4 of the hVDR cDNA, following standard colony hybridisation techniques. DNA probes were labelled by nick translation (Life Technologies, Gaithersburg, MD) with [α^{32} P] dCTP. Positively hybridising colonies were picked and secondary and tertiary screens carried out until complete purification. Cosmid DNA from positive clones was purified (Qiagen), digested with different restriction enzymes and characterised by Southern blot analysis using specific [γ^{32} P]ATP labelled oligonucleotides as probes. Cosmid clones were directly sequenced using dye-termination chemistry and automated fluorescent sequencing on an ABI Prism, 377 DNA Sequencer (Perkin-Elmer, Foster City, Ca). Sequence upstream of the most 5' cosmid was obtained by anchored PCR from genomic DNA using commercially available anchor ligated DNA (Clontech, Palo Alto, Ca).

Rapid Amplification of cDNA 5-prime Ends (5'-RACE)

Alternative 5' variants of the human VDR gene were identified by 5'RACE using commercially prepared anchor-ligated cDNA (Clontech) following the instructions of the manufacturer. Two rounds of PCR using nested reverse primers in exons 3 and 2 (P 1: 5'ccgcttcatgcttcgcctgaagaagcc-3', P2: 5'-tgcagaattcacaggtcatagcattgaag-3') were carried out on a Corbett FTS-4000 Capillary Thermal Sequencer (Corbett Research, NSW, Australia). After 26 cycles of PCR, 2% of the primary reaction was reamplified for 31 cycles. The PCR products were cloned into PUC18 and sequenced by the dideoxy chain termination method.

Cell-Culture

The embryonal kidney cell line, HEK-293, an embryonic intestine cell line, Intestine-407 and WS 1, a foetal skin fibroblast cell line were all

cultured in Eagle's MEM with Earle's BSS and supplemented with either 10% heat-inactivated FBS, 15% FBS or 10% FBS with non-essential amino acids, respectively. The osteosarcoma cell lines MG-63 and Saos-2 were cultured in Eagle's MEM with nonessential amino acids and 10% heat-inactivated FBS and McCoy's 5a medium with 15% FBS, respectively. The breast carcinoma cell line T47D and the colon carcinoma cell lines LIM 1863 and COLO 206F were cultured in RPMI medium supplemented with 0.2 IU bovine insulin/ml and 10% FBS, 5% FBS or 10% FBS, respectively. LIM 1863 were a gift from R.H. Whitehead (3). HK-2 kidney proximal tubular cells were grown in keratinocyte-serum free medium supplemented with 5ng/ml recombinant EGF, 40ug/ml bovine pituitary extract. BC1 primary foetal osteoblast-like cells were kindly donated by R. Mason (4) and were grown in Eagle's MEM with 5% FBS and 5mg/L vitamin C. Unless otherwise stated all cell lines were obtained from the American Type Culture Collection (Rockville, MD).

Reverse Transcriptase-PCR

Total RNA was extracted from approx. $1.5x10^7$ cells or leukocytes prepared from 40ml blood using guanidium isothiocyanate-cesium chloride step gradient or from human tissue using acid-phenol extraction followed by guanidium isothiocyanate-cesium chloride step gradient. First strand cDNA was synthesised from 5μg of total RNA primed with random hexamers (Promega, Madison, Wi) using Superscript II reverse transcriptase according to the manufacturers instructions (Life Technologies, Gaithersburg, MD). One-tenth of the cDNA (2μl) was used for subsequent PCR with exon 1a, 1b or 1e(5') specific forward primers (P1a, corresponding to nucleotides 1 to 21 of hVDR cDNA (1); P1b, P1e; Fig. 3) and a common reverse primer in exon 3 (corresponding to nucleotides 301 to 280 of hVDR cDNA (1)). PCR products were separated on 2% agarose and visualised with ethidium bromide staining.

RESULTS

Identification of Alternative 5' Variants of the hVDR Gene

Using 5'RACE, novel upstream exons were identified in human kidney VDR transcripts: exons 1e(5'), 1e(3'), 1b and 1c (Fig. 1). To verify these 5'RACE results and to thoroughly characterise the structure of the 5' end of

the VDR gene, exon specific forward primers were used to amplify specific VDR transcripts from human kidney RNA. These were subsequently cloned and sequenced these PCR products. Figure 1 summarises the results. Four different VDR transcripts originating from exon 1 a were identified. The major transcript (transcript 1 in figure 1) corresponds to the published cDNA sequence. Three less abundant forms (2, 3, 4 in figure 1) arise from alternative splicing of exon 1 and a 122bp exon 1c (Fig. 3a) into or out of the final transcript. These variant transcripts were recently described in a paper by Pike et. al. (2). (The exons which we have called 1 and 1c are denoted 1c and 1b respectively in this paper). Four transcripts were characterised which originated from exon 1e (5'), a novel 207bp exon >10kb upstream from exon 1a (Fig. 3b). The major 1e(5') containing transcript (9 in figure 1) consists of exon 1e(5') spliced immediately adjacent to exon 1 of the hVDR cDNA. Three less abundant variants (10, 11, 12 in figure 1) arise from alternative splicing of exon 1 and a novel 159bp exon 1e(3') (Fig. 3c) into or out of the final transcript. All these hVDR variants differ only in their 5'UTRs and potentially encode for identical proteins from translation initiation in exon 2.

Another four hVDR transcripts were identified which originate from exon 1b, a novel 96bp exon located 296bp downstream from exon 1a (Fig. 3d). The major exon 1b containing transcript (5 in figure 1) utilises exon 1b in place of exon 1a of the VDR cDNA. Two minor variants (6,7 in figure 1) arise from alternative splicing of exons 1c and 1 into or out of the transcript analogous to the exon 1a containing variants 2 and 4. Two of these exon 1b containing hVDR transcripts have the potential to encode an N-terminally variant form of the hVDR protein. Utilisation of an ATG codon in exon 1b (Fig. 3d) which is in a favourable context and in frame with the major translation start site in exon 2 could generate a protein with an additional 50 amino acids N-terminal to the ATG codon in exon 2 in the case of variant 5 or 23 amino acids in the case of variant 7. A fourth minor variant transcript containing exon 1b lacked exons 1c and 1, but included an extra 152bp of intronic sequence immediately adjacent and 3' to exon 2, potentially encoding a truncated protein as a result of an in-frame termination codon in intron 2. RT-PCR results, using exon-specific forward primers and a reverse primer in intron 2, suggest that a cryptic splice site in intron 2 is used at a low frequency in both 1a and 1b containing transcripts. A 1a containing transcript analogous to variant 8 (Fig. 1) is hypothesised (data not shown).

We hypothesize the existence of a further transcript containing exons 1b and 1c, but excluding exon 1, analogous to the 1a containing variant 3 (Fig. l).

Exon-Intron Organisation of the hVDR gene

Four overlapping cosmid clones were isolated from a human lymphocyte genomic library and characterised by hybridisation to exon-specific oligonucleotide probes. J5 extends from the 5' flanking region to intron 2; AE from intron 1c to intron 5; D2 from intron 3 to the 3'UTR; and WE from intron 6 to the 3' flanking region. The exon-intron boundaries of the hVDR gene were determined by comparison of the genomic sequence from cosmid clones with the cDNA sequence. Novel upstream exons were located in the VDR gene by sequencing cosmid clone J5. Cosmid J5 extends approx. 7kb into the intron between exons 1e(5') and 1e(3') but does not encompass exon 1e(5'). Sequence upstream of exon 1e(5') was obtained from anchored PCR clones. The hVDR spans more than 55kb and consists of 14 exons (Fig. 1). The translation start site is in exon 2 and the translation stop codon is in exon 9. Exons la and lc are composed only of 5'UTR sequence. Exons 1b and 1 have potential to encode an N-terminally variant protein.

Tissue-Specific Expression of VDR Transcripts

The pattern of expression of VDR transcripts was examined by RT-PCR in a variety of cell lines and tissues, with exon 1a, 1b or 1e specific forward primers and a common reverse primer in exon 3. Using an exon 1a specific primer, PCR products of 301, 423, 342 and 220bp, corresponding to the scheme in figure 1 (variants 1-4), were observed for all RNA samples analysed (Fig. 2a). Similarly, using an exon 1b specific primer PCR products of 305, 427, 224 and 376bp, corresponding to the scheme in figure 1 (variants 5-8), were observed for all RNA samples tested (Fig. 2b). When exon 1e(5') containing transcripts were amplified using an exon 1e(5') specific primer PCR products were only detected in RNA from human kidney tissue (two separate samples), human parathyroid adenoma tissue and an intestinal cell-line, LIM 1863 (colon carcinoma, ileococal valve) (Fig. 2c). These PCR products of 309, 468, 387 and 228bp correspond to the scheme in figure 1 (variants 9-12). RNA samples in which we were unable to detect VDR transcripts containing exon 1a were also negative for exon 1b and 1e transcripts (data not shown).

DISCUSSION

The VDR gene consists of 14 exons and spans more than 55kb of genomic DNA. It gives rise to a 427 amino-acid protein which has a domain structure common to members of the nuclear receptor superfamily.

Using 5' RACE, the present inventors have identified 5' variant transcripts of the hVDR which suggest the existence of alternative promotors. These transcripts may not have been discriminated in previous Northern analyses due to their similarity in size. Transcription initiation from exons 1a or 1e(5'), and alternative splicing, generate VDR transcripts which vary in their 5'UTRs but have the potential to encode the same 427 amino-acid protein. Transcription initiation from exon 1b, and alternative splicing, generate hVDR transcripts with the potential to encode a variant protein with an additional 50 or 23 amino-acids at the N-terminal. Although the existence of VDR isoforms has been speculated upon to account for the varied actions of vitamin D in a wide range of target tissues, the only evidence documented so far for isoforms of the human VDR is the identification of a common polymorphism in the triplet encoding the initiating methionine of the 427 amino acid form of the VDR which results in initiation of translation at an alternative start codon beginning at the 10th nucleotide downstream and potentially encodes a protein truncated by 3 amino acids at the N-terminal (5). Other evidence for VDR isoforms comes from avian species where two forms of the VDR, differing in size by 14 amino acids, are generated from a single transcript by alternative initiation of translation (6), and rat where a dominant negative VDR generated by intron retention has been described (7). 5' heterogeneity is a common feature of nuclear receptor genes and the generation of N-terminally variant protein isoforms has been described for the progesterone receptor (PR) and peroxisome proliferator-activated receptor (PPARγ) for example. Two promotors direct the expression of human PR transcripts which vary at their 5' ends and generate the two N-terminally variant isoforms, A and B, which exhibit different promotor specificities (8). Similarly two isoforms of the mouse and human PPARγ, γ1 and γ2 which contains an extra 30 amino acid N-terminal, arise from differential promotor usage and alternative splicing (9,10). Transactivation/repression of target gene expression by nuclear receptors requires multiple protein interactions

with cofactors, heterodimerisation partners and basal transcription machinery. The N-terminal domains of nuclear receptors have been shown to have transactivation function and to play a role in promotor selection. An N-terminal variant VDR protein might therefore exhibit different transactivation potential. Another possibility is that the N-terminal domain of the variant protein determines the subcellular location of the VDR protein. This would have implications for the ability of the VDR to act via non-genomic pathways.

The results demonstrate that the expression of exon 1e(5') containing transcripts is tissue-specific. RT-PCR products were only detected in kidney tissue, parathyroid adenoma tissue and an intestinal cell-line, LIM 1863. It is interesting that these tissues represent the major target tissues for the calcitropic effects of vitamin D. The absence of 1e(5') containing transcripts in two other kidney cell-lines, HK-2 (proximal tubular) and HEK-293 (embryonal kidney), as well as one other embryonic intestinal cell-line, Intestine-407, suggests that the expression of 1e transcripts may be cell-type specific.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this twenty sixth day of September 1997

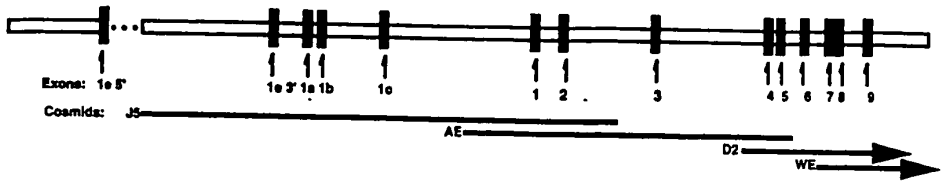
GARVAN INSTITUTE OF MEDICAL RESEARCH
Patent Attorneys for the Applicant:

F.B. RICE & CO.

References:-

1. Baker, A.R. et.al. (1988) Proc. Natl. Acad. Sci. USA 85, 3294-3298
2. Miyamoto, K. et.al. (1997) Mol. Endocrinol. 11, 1165-1179
3. Whitehead, R.H.et.al. (1987) Cancer Res. 47, 2683-2689
4. Slater, M. et al. (1994) Am. J. Physiology 267, E990-1001
5. Saijo, T.et.al. (1991) Am.J.Hum.Genet. 49, 668-673
6. Lu, Z. et.al. (1997) Arch.Biochem.Biophys. 339, 99-106
7. Ebihara, K. et.al. (1996) Mol.Cell.Biol. 16, 3393-3400
8. Kastner, P. et.al. (1990) EMBO J. 9, 1603-1614
9. Zhu, Y. et.al. (1995) Proc.Natl.Acad.Sci.USA 92, 7921-7925
10. Mukherjee, R. et.al. (1997) J.Biol.Chem. 272, 8071-8076

A.



B.

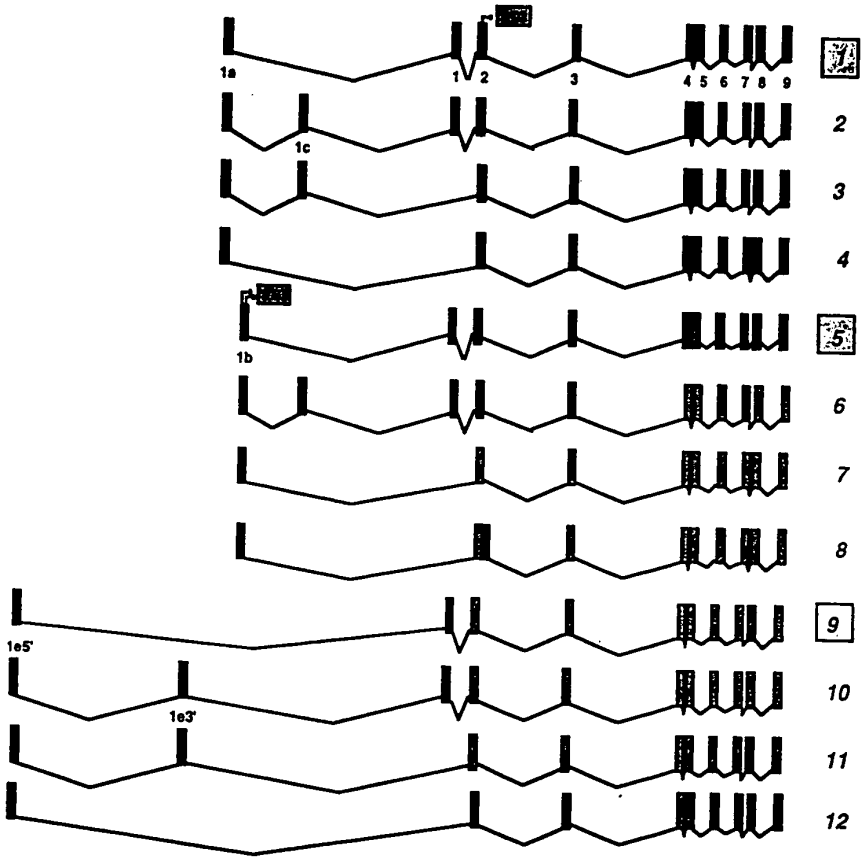


Figure 1. Human vitamin D receptor gene locus.

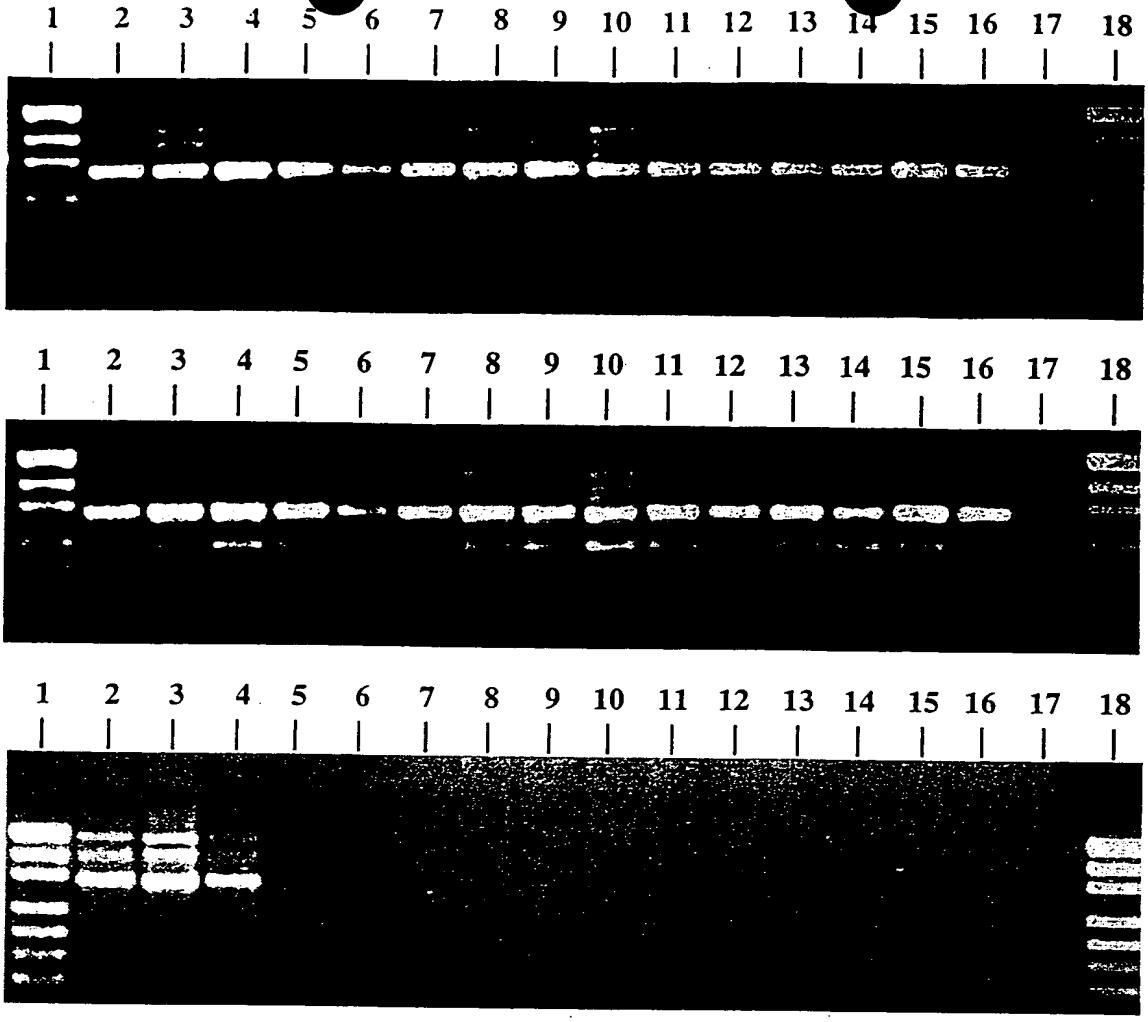


Figure 2. RT-PCR analysis of hVDR gene expression.

A. 5'...atcccttaag GGCTCCTGAACCTAGCCCAGCTGGACGGAG
AAATGGACTCTAGCCTCCTCTGATAGCCTCATGCCAGGCCC
CGTGCACATTGCTTTGCTTGCCTCCCTCAATCCTCATAGCT
TCTCTTTGGGgtaagtacag...3'

B. 5'...TGCGACCTTGGCGGTGAGCCTGGGGACAGGGGTGAGGC
CAGAGACGGACGGACGCAGGGGCCCGGCCCAAGGCGAGGG
AGAACAGCGGCACTAAGGCAGAAAGGAAGAGGGCGGTGTG
TTCACCCGCAGCCCAATCCATCACTCAGCAACTCCTAGAC
GCTGGTAGAAAGTTCCTCCGAGGAGCCTGCCATCCAGTCGT
GCGTGCAG...3'

C. 5'...tgttttttag AGGCAGCATGAAACAGTGGGATGTGCAGAG
AGAAGATCTGGGTCCAGTAGCTCTGACACTCCTCAGCTGT
AGAAACCTTGACAACTCTGCACATCAGTTGTACAATGGAA
CGGTATTTTTTACTCTTCATGTCTGAAAAGGCTATGATAA
AGATCAAgtaagatatt...3'

D. 5'...GTTTCCTTCTTCTGTCGGGGCGCCTTGGCATGGAGTGG
AGGAATAAGAAAAGGAGCGATTGGCTGTCGATGGTGCTCA
GAACTGCTGGAGTGGAGGgtgtgtaacc...3'

Figure 3. Nucleotide sequence of novel exons detected by 5'RACE.

Figure 4. Transcript 5:

-nucleotides 1 - 96 correspond to exon 1b
-nucleotides 97 - 1463 correspond to exons 1 to the stop codon in exon 9 (or nucleotides -83 - 1283 of the hVDR cDNA (1))

(Sequence Range: 1 to 1463)

```
          10         20         30         40         50
     *    *     *    *     *    *     *    *     *    *
GTTTCCTTCT TCTGTCGGGG CGCCTTGGCA TGGAGTGGAG GAATAAGAAA
CAAAGGAAGA AGACAGCCCC GCGGAACCGT ACCTCACCTC CTTATTCTTT
                                 MetGluTrpArg AsnLysLys>

          60         70         80         90         100
     *    *     *    *     *    *     *    *     *    *
AGGAGCGATT GGCTGTCGAT GGTGCTCAGA ACTGCTGGAG TGGAGGAAGC
TCCTCGCTAA CCGACAGCTA CCACGAGTCT TGACGACCTC ACCTCCTTCG
ArgSerAsp TrpLeuSerMet ValLeuArg ThrAlaGly ValGluGluAla>

        110        120        130        140        150
     *    *     *    *     *    *     *    *     *    *
CTTTGGGTCT GAAGTGTCTG TGAGACCTCA CAGAAGAGCA CCCCTGGGCT
GAAACCCAGA CTTCACAGAC ACTCTGGAGT GTCTTCTCGT GGGGACCCGA
 PheGlySer GluValSer ValArgProHis ArgArgAla ProLeuGly>

        160        170        180        190        200
     *    *     *    *     *    *     *    *     *    *
CCACTTACCT GCCCCCTGCT CCTTCAGGGA TGGAGGCAAT GGCGGCCAGC
GGTGAATGGA CGGGGGACGA GGAAGTCCCT ACCTCCGTTA CCGCCGGTCG
SerThrTyrLeu ProProAla ProSerGly MetGluAlaMet AlaAlaSer>

        210        220        230        240        250
     *    *     *    *     *    *     *    *     *    *
ACTTCCCTGC CTGACCCTGG AGACTTTGAC CGGAACGTGC CCCGGATCTG
TGAAGGGACG GACTGGGACC TCTGAAACTG GCCTTGCACG GGGCCTAGAC
ThrSerLeu ProAspProGly AspPheAsp ArgAsnVal ProArgIleCys>

        260        270        280        290        300
     *    *     *    *     *    *     *    *     *    *
TGGGGTGTGT GGAGACCGAG CCACTGGCTT TCACTTCAAT GCTATGACCT
ACCCCACACA CCTCTGGCTC GGTGACCGAA AGTGAAGTTA CGATACTGGA
 GlyValCys GlyAspArg AlaThrGlyPhe HisPheAsn AlaMetThr>

        310        320        330        340        350
     *    *     *    *     *    *     *    *     *    *
GTGAAGGCTG CAAAGGCTTC TTCAGGCGAA GCATGAAGCG GAAGGCACTA
CACTTCCGAC GTTTCCGAAG AAGTCCGCTT CGTACTTCGC CTTCCGTGAT
CysGluGlyCys LysGlyPhe PheArgArg SerMetLysArg LysAlaLeu>

        360        370        380        390        400
     *    *     *    *     *    *     *    *     *    *
TTCACCTGCC CCTTCAACGG GGACTGCCGC ATCACCAAGG ACAACCGACG
AAGTGGACGG GGAAGTTGCC CCTGACGGCG TAGTGGTTCC TGTTGGCTGC
PheThrCys ProPheAsnGly AspCysArg IleThrLys AspAsnArgArg>
```

```
       410        420        430        440        450
    *   *      *   *      *   *      *   *      *   *
CCACTGCCAG GCCTGCCGGC TCAAACGCTG TGTGGACATC GGCATGATGA
GGTGACGGTC CGGACGGCCG AGTTTGCGAC ACACCTGTAG CCGTACTACT
 HisCysGln AlaCysArg LeuLysArgCys ValAspIle GlyMetMet>

       460        470        480        490        500
    *   *      *   *      *   *      *   *      *   *
AGGAGTTCAT TCTGACAGAT GAGGAAGTGC AGAGGAAGCG GGAGATGATC
TCCTCAAGTA AGACTGTCTA CTCCTTCACG TCTCCTTCGC CCTCTACTAG
LysGluPheIle LeuThrAsp GluGluVal GlnArgLysArg GluMetIle>

       510        520        530        540        550
    *   *      *   *      *   *      *   *      *   *
CTGAAGCGGA AGGAGGAGGA GGCCTTGAAG GACAGTCTGC GGCCCAAGCT
GACTTCGCCT TCCTCCTCCT CCGGAACTTC CTGTCAGACG CCGGGTTCGA
LeuLysArg LysGluGluGlu AlaLeuLys AspSerLeu ArgProLysLeu>

       560        570        580        590        600
    *   *      *   *      *   *      *   *      *   *
GTCTGAGGAG CAGCAGCGCA TCATTGCCAT ACTGCTGGAC GCCCACCATA
CAGACTCCTC GTCGTCGCGT AGTAACGGTA TGACGACCTG CGGGTGGTAT
 SerGluGlu GlnGlnArg IleIleAlaIle LeuLeuAsp AlaHisHis>

       610        620        630        640        650
    *   *      *   *      *   *      *   *      *   *
AGACCTACGA CCCCACCTAC TCCGACTTCT GCCAGTTCCG GCCTCCAGTT
TCTGGATGCT GGGGTGGATG AGGCTGAAGA CGGTCAAGGC CGGAGGTCAA
LysThrTyrAsp ProThrTyr SerAspPhe CysGlnPheArg ProProVal>

       660        670        680        690        700
    *   *      *   *      *   *      *   *      *   *
CGTGTGAATG ATGGTGGAGG GAGCCATCCT TCCAGGCCCA ACTCCAGACA
GCACACTTAC TACCACCTCC CTCGGTAGGA AGGTCCGGGT TGAGGTCTGT
ArgValAsn AspGlyGlyGly SerHisPro SerArgPro AsnSerArgHis>

       710        720        730        740        750
    *   *      *   *      *   *      *   *      *   *
CACTCCCAGC TTCTCTGGGG ACTCCTCCTC CTCCTGCTCA GATCACTGTA
GTGAGGGTCG AAGAGACCCC TGAGGAGGAG GAGGACGAGT CTAGTGACAT
 ThrProSer PheSerGly AspSerSerSer SerCysSer AspHisCys>

       760        770        780        790        800
    *   *      *   *      *   *      *   *      *   *
TCACCTCTTC AGACATGATG GACTCGTCCA GCTTCTCCAA TCTGGATCTG
AGTGGAGAAG TCTGTACTAC CTGAGCAGGT CGAAGAGGTT AGACCTAGAC
IleThrSerSer AspMetMet AspSerSer SerPheSerAsn LeuAspLeu>

       810        820        830        840        850
    *   *      *   *      *   *      *   *      *   *
AGTGAAGAAG ATTCAGATGA CCCTTCTGTG ACCCTAGAGC TGTCCCAGCT
TCACTTCTTC TAAGTCTACT GGGAAGACAC TGGGATCTCG ACAGGGTCGA
SerGluGlu AspSerAspAsp ProSerVal ThrLeuGlu LeuSerGlnLeu>
```

```
       860        870        880        890        900
    *    *     *    *     *    *     *    *     *    *
CTCCATGCTG CCCCACCTGG CTGACCTGGT CAGTTACAGC ATCCAAAAGG
GAGGTACGAC GGGGTGGACC GACTGGACCA GTCAATGTCG TAGGTTTTCC
 SerMetLeu ProHisLeu AlaAspLeuVal SerTyrSer IleGlnLys>

       910        920        930        940        950
    *    *     *    *     *    *     *    *     *    *
TCATTGGCTT TGCTAAGATG ATACCAGGAT TCAGAGACCT CACCTCTGAG
AGTAACCGAA ACGATTCTAC TATGGTCCTA AGTCTCTGGA GTGGAGACTC
ValIleGlyPhe AlaLysMet IleProGly PheArgAspLeu ThrSerGlu>

       960        970        980        990       1000
    *    *     *    *     *    *     *    *     *    *
GACCAGATCG TACTGCTGAA GTCAAGTGCC ATTGAGGTCA TCATGTTGCG
CTGGTCTAGC ATGACGACTT CAGTTCACGG TAACTCCAGT AGTACAACGC
AspGlnIle ValLeuLeuLys SerSerAla IleGluVal IleMetLeuArg>

      1010       1020       1030       1040       1050
    *    *     *    *     *    *     *    *     *    *
CTCCAATGAG TCCTTCACCA TGGACGACAT GTCCTGGACC TGTGGCAACC
GAGGTTACTC AGGAAGTGGT ACCTGCTGTA CAGGACCTGG ACACCGTTGG
 SerAsnGlu SerPheThr MetAspAspMet SerTrpThr CysGlyAsn>

      1060       1070       1080       1090       1100
    *    *     *    *     *    *     *    *     *    *
AAGACTACAA GTACCGCGTC AGTGACGTGA CCAAAGCCGG ACACAGCCTG
TTCTGATGTT CATGGCGCAG TCACTGCACT GGTTTCGGCC TGTGTCGGAC
GlnAspTyrLys TyrArgVal SerAspVal ThrLysAlaGly HisSerLeu>

      1110       1120       1130       1140       1150
    *    *     *    *     *    *     *    *     *    *
GAGCTGATTG AGCCCCTCAT CAAGTTCCAG GTGGGACTGA AGAAGCTGAA
CTCGACTAAC TCGGGGAGTA GTTCAAGGTC CACCCTGACT TCTTCGACTT
GluLeuIle GluProLeuIle LysPheGln ValGlyLeu LysLysLeuAsn>

      1160       1170       1180       1190       1200
    *    *     *    *     *    *     *    *     *    *
CTTGCATGAG GAGGAGCATG TCCTGCTCAT GGCCATCTGC ATCGTCTCCC
GAACGTACTC CTCCTCGTAC AGGACGAGTA CCGGTAGACG TAGCAGAGGG
 LeuHisGlu GluGluHis ValLeuLeuMet AlaIleCys IleValSer>

      1210       1220       1230       1240       1250
    *    *     *    *     *    *     *    *     *    *
CAGATCGTCC TGGGGTGCAG GACGCCGCGC TGATTGAGGC CATCCAGGAC
GTCTAGCAGG ACCCCACGTC CTGCGGCGCG ACTAACTCCG GTAGGTCCTG
ProAspArgPro GlyValGln AspAlaAla LeuIleGluAla IleGlnAsp>

      1260       1270       1280       1290       1300
    *    *     *    *     *    *     *    *     *    *
CGCCTGTCCA ACACACTGCA GACGTACATC CGCTGCCGCC ACCCGCCCCC
GCGGACAGGT TGTGTGACGT CTGCATGTAG GCGACGGCGG TGGGCGGGGG
ArgLeuSer AsnThrLeuGln ThrTyrIle ArgCysArg HisProProPro>
```

```
      1310       1320       1330       1340       1350
    *    *     *    *     *    *     *    *     *    *
GGGCAGCCAC CTGCTCTATG CCAAGATGAT CCAGAAGCTA GCCGACCTGC
CCCGTCGGTG GACGAGATAC GGTTCTACTA GGTCTTCGAT CGGCTGGACG
 GlySerHis LeuLeuTyr AlaLysMetIle GlnLysLeu AlaAspLeu>

      1360       1370       1380       1390       1400
    *    *     *    *     *    *     *    *     *    *
GCAGCCTCAA TGAGGAGCAC TCCAAGCAGT ACCGCTGCCT CTCCTTCCAG
CGTCGGAGTT ACTCCTCGTG AGGTTCGTCA TGGCGACGGA GAGGAAGGTC
ArgSerLeuAsn GluGluHis SerLysGln TyrArgCysLeu SerPheGln>

      1410       1420       1430       1440       1450
    *    *     *    *     *    *     *    *     *    *
CCTGAGTGCA GCATGAAGCT AACGCCCCTT GTGCTCGAAG TGTTTGGCAA
GGACTCACGT CGTACTTCGA TTGCGGGGAA CACGAGCTTC ACAAACCGTT
ProGluCys SerMetLysLeu ThrProLeu ValLeuGlu ValPheGlyAsn>

      1460
    *    *
TGAGATCTCC TGA
ACTCTAGAGG ACT
 GluIleSer ***>
```

Figure 5. Transcript 7:

-nucleotides 1 - 96 correspond to exon 1b
-nucleotides 97 - 1382 correspond to exons 2 to the stop codon in exon 9 (or nucleotides -2 - 1283 of the hVDR cDNA sequence(1)).

(Sequence Range: 1 to 1382)

```
          10         20         30         40         50
      *    *     *    *     *    *     *    *     *    *
GTTTCCTTCT TCTGTCGGGG CGCCTTGGCA TGGAGTGGAG GAATAAGAAA
CAAAGGAAGA AGACAGCCCC GCGGAACCGT ACCTCACCTC CTTATTCTTT
                                 MetGluTrpArg AsnLysLys>

          60         70         80         90        100
      *    *     *    *     *    *     *    *     *    *
AGGAGCGATT GGCTGTCGAT GGTGCTCAGA ACTGCTGGAG TGGAGGGGAT
TCCTCGCTAA CCGACAGCTA CCACGAGTCT TGACGACCTC ACCTCCCCTA
ArgSerAsp TrpLeuSerMet ValLeuArg ThrAlaGly ValGluGlyMet>

         110        120        130        140        150
      *    *     *    *     *    *     *    *     *    *
GGAGGCAATG GCGGCCAGCA CTTCCCTGCC TGACCCTGGA GACTTTGACC
CCTCCGTTAC CGCCGGTCGT GAAGGGACGG ACTGGGACCT CTGAAACTGG
 GluAlaMet AlaAlaSer ThrSerLeuPro AspProGly AspPheAsp>

         160        170        180        190        200
      *    *     *    *     *    *     *    *     *    *
GGAACGTGCC CCGGATCTGT GGGGTGTGTG GAGACCGAGC CACTGGCTTT
CCTTGCACGG GGCCTAGACA CCCCACACAC CTCTGGCTCG GTGACCGAAA
ArgAsnValPro ArgIleCys GlyValCys GlyAspArgAla ThrGlyPhe>

         210        220        230        240        250
      *    *     *    *     *    *     *    *     *    *
CACTTCAATG CTATGACCTG TGAAGGCTGC AAAGGCTTCT TCAGGCGAAG
GTGAAGTTAC GATACTGGAC ACTTCCGACG TTTCCGAAGA AGTCCGCTTC
HisPheAsn AlaMetThrCys GluGlyCys LysGlyPhe PheArgArgSer>

         260        270        280        290        300
      *    *     *    *     *    *     *    *     *    *
CATGAAGCGG AAGGCACTAT TCACCTGCCC CTTCAACGGG GACTGCCGCA
GTACTTCGCC TTCCGTGATA AGTGGACGGG GAAGTTGCCC CTGACGGCGT
 MetLysArg LysAlaLeu PheThrCysPro PheAsnGly AspCysArg>

         310        320        330        340        350
      *    *     *    *     *    *     *    *     *    *
TCACCAAGGA CAACCGACGC CACTGCCAGG CCTGCCGGCT CAAACGCTGT
AGTGGTTCCT GTTGGCTGCG GTGACGGTCC GGACGGCCGA GTTTGCGACA
IleThrLysAsp AsnArgArg HisCysGln AlaCysArgLeu LysArgCys>

         360        370        380        390        400
      *    *     *    *     *    *     *    *     *    *
GTGGACATCG GCATGATGAA GGAGTTCATT CTGACAGATG AGGAAGTGCA
CACCTGTAGC CGTACTACTT CCTCAAGTAA GACTGTCTAC TCCTTCACGT
ValAspIle GlyMetMetLys GluPheIle LeuThrAsp GluGluValGln>
```

```
       410        420        430        440        450
    *    *     *    *     *    *     *    *     *    *
GAGGAAGCGG GAGATGATCC TGAAGCGGAA GGAGGAGGAG GCCTTGAAGG
CTCCTTCGCC CTCTACTAGG ACTTCGCCTT CCTCCTCCTC CGGAACTTCC
 ArgLysArg GluMetIle LeuLysArgLys GluGluGlu AlaLeuLys>

       460        470        480        490        500
    *    *     *    *     *    *     *    *     *    *
ACAGTCTGCG GCCCAAGCTG TCTGAGGAGC AGCAGCGCAT CATTGCCATA
TGTCAGACGC CGGGTTCGAC AGACTCCTCG TCGTCGCGTA GTAACGGTAT
AspSerLeuArg ProLysLeu SerGluGlu GlnGlnArgIle IleAlaIle>

       510        520        530        540        550
    *    *     *    *     *    *     *    *     *    *
CTGCTGGACG CCCACCATAA GACCTACGAC CCCACCTACT CCGACTTCTG
GACGACCTGC GGGTGGTATT CTGGATGCTG GGGTGGATGA GGCTGAAGAC
LeuLeuAsp AlaHisHisLys ThrTyrAsp ProThrTyr SerAspPheCys>

       560        570        580        590        600
    *    *     *    *     *    *     *    *     *    *
CCAGTTCCGG CCTCCAGTTC GTGTGAATGA TGGTGGAGGG AGCCATCCTT
GGTCAAGGCC GGAGGTCAAG CACACTTACT ACCACCTCCC TCGGTAGGAA
 GlnPheArg ProProVal ArgValAsnAsp GlyGlyGly SerHisPro>

       610        620        630        640        650
    *    *     *    *     *    *     *    *     *    *
CCAGGCCCAA CTCCAGACAC ACTCCCAGCT TCTCTGGGGA CTCCTCCTCC
GGTCCGGGTT GAGGTCTGTG TGAGGGTCGA AGAGACCCCT GAGGAGGAGG
SerArgProAsn SerArgHis ThrProSer PheSerGlyAsp SerSerSer>

       660        670        680        690        700
    *    *     *    *     *    *     *    *     *    *
TCCTGCTCAG ATCACTGTAT CACCTCTTCA GACATGATGG ACTCGTCCAG
AGGACGAGTC TAGTGACATA GTGGAGAAGT CTGTACTACC TGAGCAGGTC
SerCysSer AspHisCysIle ThrSerSer AspMetMet AspSerSerSer>

       710        720        730        740        750
    *    *     *    *     *    *     *    *     *    *
CTTCTCCAAT CTGGATCTGA GTGAAGAAGA TTCAGATGAC CCTTCTGTGA
GAAGAGGTTA GACCTAGACT CACTTCTTCT AAGTCTACTG GGAAGACACT
 PheSerAsn LeuAspLeu SerGluGluAsp SerAspAsp ProSerVal>

       760        770        780        790        800
    *    *     *    *     *    *     *    *     *    *
CCCTAGAGCT GTCCCAGCTC TCCATGCTGC CCCACCTGGC TGACCTGGTC
GGGATCTCGA CAGGGTCGAG AGGTACGACG GGGTGGACCG ACTGGACCAG
ThrLeuGluLeu SerGlnLeu SerMetLeu ProHisLeuAla AspLeuVal>

       810        820        830        840        850
    *    *     *    *     *    *     *    *     *    *
AGTTACAGCA TCCAAAAGGT CATTGGCTTT GCTAAGATGA TACCAGGATT
TCAATGTCGT AGGTTTTCCA GTAACCGAAA CGATTCTACT ATGGTCCTAA
SerTyrSer IleGlnLysVal IleGlyPhe AlaLysMet IleProGlyPhe>
```

```
       860        870        880        890        900
    *   *    *   *    *   *    *   *    *   *
CAGAGACCTC ACCTCTGAGG ACCAGATCGT ACTGCTGAAG TCAAGTGCCA
GTCTCTGGAG TGGAGACTCC TGGTCTAGCA TGACGACTTC AGTTCACGGT
 ArgAspLeu ThrSerGlu AspGlnIleVal LeuLeuLys SerSerAla>

       910        920        930        940        950
    *   *    *   *    *   *    *   *    *   *
TTGAGGTCAT CATGTTGCGC TCCAATGAGT CCTTCACCAT GGACGACATG
AACTCCAGTA GTACAACGCG AGGTTACTCA GGAAGTGGTA CCTGCTGTAC
IleGluValIle MetLeuArg SerAsnGlu SerPheThrMet AspAspMet>

       960        970        980        990       1000
    *   *    *   *    *   *    *   *    *   *
TCCTGGACCT GTGGCAACCA AGACTACAAG TACCGCGTCA GTGACGTGAC
AGGACCTGGA CACCGTTGGT TCTGATGTTC ATGGCGCAGT CACTGCACTG
SerTrpThr CysGlyAsnGln AspTyrLys TyrArgVal SerAspValThr>

      1010       1020       1030       1040       1050
    *   *    *   *    *   *    *   *    *   *
CAAAGCCGGA CACAGCCTGG AGCTGATTGA GCCCCTCATC AAGTTCCAGG
GTTTCGGCCT GTGTCGGACC TCGACTAACT CGGGGAGTAG TTCAAGGTCC
 LysAlaGly HisSerLeu GluLeuIleGlu ProLeuIle LysPheGln>

      1060       1070       1080       1090       1100
    *   *    *   *    *   *    *   *    *   *
TGGGACTGAA GAAGCTGAAC TTGCATGAGG AGGAGCATGT CCTGCTCATG
ACCCTGACTT CTTCGACTTG AACGTACTCC TCCTCGTACA GGACGAGTAC
ValGlyLeuLys LysLeuAsn LeuHisGlu GluGluHisVal LeuLeuMet>

      1110       1120       1130       1140       1150
    *   *    *   *    *   *    *   *    *   *
GCCATCTGCA TCGTCTCCCC AGATCGTCCT GGGGTGCAGG ACGCCGCGCT
CGGTAGACGT AGCAGAGGGG TCTAGCAGGA CCCCACGTCC TGCGGCGCGA
AlaIleCys IleValSerPro AspArgPro GlyValGln AspAlaAlaLeu>

      1160       1170       1180       1190       1200
    *   *    *   *    *   *    *   *    *   *
GATTGAGGCC ATCCAGGACC GCCTGTCCAA CACACTGCAG ACGTACATCC
CTAACTCCGG TAGGTCCTGG CGGACAGGTT GTGTGACGTC TGCATGTAGG
 IleGluAla IleGlnAsp ArgLeuSerAsn ThrLeuGln ThrTyrIle>

      1210       1220       1230       1240       1250
    *   *    *   *    *   *    *   *    *   *
GCTGCCGCCA CCCGCCCCCG GGCAGCCACC TGCTCTATGC CAAGATGATC
CGACGGCGGT GGGCGGGGGC CCGTCGGTGG ACGAGATACG GTTCTACTAG
ArgCysArgHis ProProPro GlySerHis LeuLeuTyrAla LysMetIle>

      1260       1270       1280       1290       1300
    *   *    *   *    *   *    *   *    *   *
CAGAAGCTAG CCGACCTGCG CAGCCTCAAT GAGGAGCACT CCAAGCAGTA
GTCTTCGATC GGCTGGACGC GTCGGAGTTA CTCCTCGTGA GGTTCGTCAT
GlnLysLeu AlaAspLeuArg SerLeuAsn GluGluHis SerLysGlnTyr>
```

```
      1310       1320       1330       1340       1350
    *    *     *    *     *    *     *    *     *    *
CCGCTGCCTC TCCTTCCAGC CTGAGTGCAG CATGAAGCTA ACGCCCCTTG
GGCGACGGAG AGGAAGGTCG GACTCACGTC GTACTTCGAT TGCGGGGAAC
 ArgCysLeu SerPheGln ProGluCysSer MetLysLeu ThrProLeu>

      1360       1370       1380
    *    *     *    *     *    *
TGCTCGAAGT GTTTGGCAAT GAGATCTCCT GA
ACGAGCTTCA CAAACCGTTA CTCTAGAGGA CT
ValLeuGluVal PheGlyAsn GluIleSer ***>
```

Figure 6. Transcript 8:

-nucleotides 1 - 96 correspond to exon 1b
-nucleotides 97 - 244 correspond to exon 2
-nucleotides 245 - 396 correspond to intronic sequence immediately 3' to exon 2
-nucleotides 397 - 1534 correspond to exons 3 to the stop codon in exon 9 (or nucleotides 146 - 1283 of the hVDR cDNA sequence (1))

(Sequence Range: 1 to 1534)

```
          10         20         30         40         50
      *    *     *    *     *    *     *    *     *    *
GTTTCCTTCT TCTGTCGGGG CGCCTTGGCA TGGAGTGGAG GAATAAGAAA
CAAAGGAAGA AGACAGCCCC GCGGAACCGT ACCTCACCTC CTTATTCTTT
                                 MetGluTrpArg AsnLysLys>

          60         70         80         90         100
      *    *     *    *     *    *     *    *     *    *
AGGAGCGATT GGCTGTCGAT GGTGCTCAGA ACTGCTGGAG TGGAGGGGAT
TCCTCGCTAA CCGACAGCTA CCACGAGTCT TGACGACCTC ACCTCCCCTA
ArgSerAsp TrpLeuSerMet ValLeuArg ThrAlaGly ValGluGlyMet>

         110        120        130        140        150
      *    *     *    *     *    *     *    *     *    *
GGAGGCAATG GCGGCCAGCA CTTCCCTGCC TGACCCTGGA GACTTTGACC
CCTCCGTTAC CGCCGGTCGT GAAGGGACGG ACTGGGACCT CTGAAACTGG
 GluAlaMet AlaAlaSer ThrSerLeuPro AspProGly AspPheAsp>

         160        170        180        190        200
      *    *     *    *     *    *     *    *     *    *
GGAACGTGCC CCGGATCTGT GGGGTGTGTG GAGACCGAGC CACTGGCTTT
CCTTGCACGG GGCCTAGACA CCCCACACAC CTCTGGCTCG GTGACCGAAA
ArgAsnValPro ArgIleCys GlyValCys GlyAspArgAla ThrGlyPhe>

         210        220        230        240        250
      *    *     *    *     *    *     *    *     *    *
CACTTCAATG CTATGACCTG TGAAGGCTGC AAAGGCTTCT TCAGGTGAGC
GTGAAGTTAC GATACTGGAC ACTTCCGACG TTTCCGAAGA AGTCCACTCG
HisPheAsn AlaMetThrCys GluGlyCys LysGlyPhe PheArg***

         260        270        280        290        300
      *    *     *    *     *    *     *    *     *    *
CCCCCTCCCA GGCTCTCCCC AGTGGAAAGG GAGGGAGAAG AAGCAAGGTG
GGGGGAGGGT CCGAGAGGGG TCACCTTTCC CTCCCTCTTC TTCGTTCCAC

         310        320        330        340        350
      *    *     *    *     *    *     *    *     *    *
TTTCCATGAA GGGAGCCCTT GCATTTTTCA CATCTCCTTC CTTACAATGT
AAAGGTACTT CCCTCGGGAA CGTAAAAAGT GTAGAGGAAG GAATGTTACA

         360        370        380        390        400
      *    *     *    *     *    *     *    *     *    *
CCATGGAACA TGCGGCGCTC ACAGCCACAG GAGCAGGAGG GTCTTGGCGA
GGTACCTTGT ACGCCGCGAG TGTCGGTGTC CTCGTCCTCC CAGAACCGCT
```

```
       410         420         430         440         450
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
AGCATGAAGC GGAAGGCACT ATTCACCTGC CCCTTCAACG GGGACTGCCG
TCGTACTTCG CCTTCCGTGA TAAGTGGACG GGGAAGTTGC CCCTGACGGC

       460         470         480         490         500
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
CATCACCAAG GACAACCGAC GCCACTGCCA GGCCTGCCGG CTCAAACGCT
GTAGTGGTTC CTGTTGGCTG CGGTGACGGT CCGGACGGCC GAGTTTGCGA

       510         520         530         540         550
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
GTGTGGACAT CGGCATGATG AAGGAGTTCA TTCTGACAGA TGAGGAAGTG
CACACCTGTA GCCGTACTAC TTCCTCAAGT AAGACTGTCT ACTCCTTCAC

       560         570         580         590         600
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
CAGAGGAAGC GGGAGATGAT CCTGAAGCGG AAGGAGGAGG AGGCCTTGAA
GTCTCCTTCG CCCTCTACTA GGACTTCGCC TTCCTCCTCC TCCGGAACTT

       610         620         630         640         650
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
GGACAGTCTG CGGCCCAAGC TGTCTGAGGA GCAGCAGCGC ATCATTGCCA
CCTGTCAGAC GCCGGGTTCG ACAGACTCCT CGTCGTCGCG TAGTAACGGT

       660         670         680         690         700
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
TACTGCTGGA CGCCCACCAT AAGACCTACG ACCCCACCTA CTCCGACTTC
ATGACGACCT GCGGGTGGTA TTCTGGATGC TGGGGTGGAT GAGGCTGAAG

       710         720         730         740         750
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
TGCCAGTTCC GGCCTCCAGT TCGTGTGAAT GATGGTGGAG GGAGCCATCC
ACGGTCAAGG CCGGAGGTCA AGCACACTTA CTACCACCTC CCTCGGTAGG

       760         770         780         790         800
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
TTCCAGGCCC AACTCCAGAC ACACTCCCAG CTTCTCTGGG GACTCCTCCT
AAGGTCCGGG TTGAGGTCTG TGTGAGGGTC GAAGAGACCC CTGAGGAGGA

       810         820         830         840         850
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
CCTCCTGCTC AGATCACTGT ATCACCTCTT CAGACATGAT GGACTCGTCC
GGAGGACGAG TCTAGTGACA TAGTGGAGAA GTCTGTACTA CCTGAGCAGG

       860         870         880         890         900
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
AGCTTCTCCA ATCTGGATCT GAGTGAAGAA GATTCAGATG ACCCTTCTGT
TCGAAGAGGT TAGACCTAGA CTCACTTCTT CTAAGTCTAC TGGGAAGACA

       910         920         930         940         950
    *   *   *   *   *   *   *   *   *   *   *   *   *   *
GACCCTAGAG CTGTCCCAGC TCTCCATGCT GCCCCACCTG GCTGACCTGG
CTGGGATCTC GACAGGGTCG AGAGGTACGA CGGGGTGGAC CGACTGGACC
```

```
      1510        1520        1530
       *     *     *     *     *     *
TGTGCTCGAA GTGTTTGGCA ATGAGATCTC CTGA
ACACGAGCTT CACAAACCGT TACTCTAGAG GACT
```